What is claimed is:

1. An apparatus for encoding a DNA sequence, which comprises:

a comparative unit aligning a reference sequence having known DNA information with a subject sequence to be encoded and extracting a difference between the reference sequence and the subject sequence;

a conversion unit converting information of the extracted difference between the reference sequence and the subject sequence into a string of predetermined characters;

a code storage unit storing predetermined conversion codes that correspond to the individual characters; and

an encoding unit encoding the individual characters that make the string of the characters using the conversion codes.

2. The apparatus of claim 1, wherein the characters comprises a first character representing DNA base symbols, a second character representing the number of the difference, a third character representing the starting and ending of the difference, and a fourth character representing continuation of the difference.

3. The apparatus of claim 2, wherein the conversion unit converts respective information of starting, start position, continuation, the number of continued bases, bases, ending, and distance between the start position and the end position of the difference into the third character, the second character, the fourth character, the second character, the first character, the third character, and the second character, and outputs the string of the characters.

4. The apparatus of claim 1, wherein the difference comprises start region mismatch between the reference sequence and the subject sequence, blank by base deletion of the subject sequence corresponding to the reference sequence, single base pair mismatch between the reference sequence and the subject sequence, base insertion into the subject sequence, multiple base pair mismatch between the reference sequence and the subject sequence, and end region mismatch between the reference sequence and the subject reference.

5. The apparatus of claim 1, wherein the conversion codes are 4 bit codes, each of which corresponds to each of the characters.

6. The apparatus of claim 1, which further comprises a division unit dividing the extracted difference into segments of predetermined sizes, and

wherein the conversion unit converts information of the extracted difference into the string of the characters based on the segments.

7. The apparatus of claim 1, which further comprises:

a compression unit compressing the encoded subject sequence; and

a sequence storage unit storing the compressed subject sequence.

8. The apparatus of claim 1, which further comprises a pre-processing unit creating a variation sequence generation factor from a variation sequence generation function that uses random variables as inputs and modifying the reference sequence using the created variation sequence generation factor.

9. The apparatus of claim 8, wherein the variation sequence induction factor comprises the total number of variations, distance between the variations, length of the variations, type of the variations, and a variation sequence.

10. A method for encoding a DNA sequence, which comprises:

aligning a reference sequence having known DNA information with a subject sequence to be encoded;

extracting a difference between the reference sequence and the subject sequence;

converting information of the extracted difference between the reference sequence and the subject sequence into a string of predetermined characters; and

encoding the individual characters that make the string of the predetermined characters using predetermined conversion codes that correspond to the individual characters.

11. The method of claim 10, wherein the characters comprises a first character representing DNA base symbols, a second character representing the

number of the difference, a third character representing the starting and ending of the difference, and a fourth character representing continuation of the difference.

12. The method of claim 11, wherein converting comprises:

allotting the third character for the starting of the difference;

allotting the second character for the starting position of the difference;

allotting the fourth character for the continuation of the difference;

allotting the second character for the number of the continued bases of the difference;

allotting the first character for the bases of the difference;

allotting the third character for the ending of the difference;

allotting the second character for the distance between the start position and the end position of the difference; and

outputting the string of the allotted characters.

13. The method of claim 10, wherein the difference comprises start region mismatch between the reference sequence and the subject sequence, blank by base deletion of the subject sequence corresponding to the reference sequence, single base pair mismatch between the reference sequence and the subject sequence, base insertion into the subject sequence, multiple base pair mismatch between the reference sequence and the subject sequence, and end region mismatch between the reference sequence and the subject reference.

14. The method of claim 10, wherein the conversion codes are 4 bit codes, each of which corresponds to each of the characters.

15. The method of claim 10, which further comprises dividing the extracted difference into segments of predetermined sizes, and

wherein in converting, information of the extracted difference is converted into the string of the characters based on the segments.

16. The method of claim 10, which further comprises:

compressing the encoded subject sequence; and

storing the compressed subject sequence.

17. The method of claim 10, which further comprises, before aligning, creating a variation sequence induction factor from a variation sequence induction function that uses random variables as inputs and modifying the reference sequence using the created variation sequence induction factor.

18. The method of claim 17, wherein the variation sequence induction factor comprises the total number of variations, distance between the variations, length of the variations, type of the variations, and a variation sequence.

19. A computer readable medium having embodied thereon a computer program for a method for encoding a DNA sequence, the method comprising:

aligning a reference sequence having known DNA information with a subject sequence to be encoded;

extracting a difference between the reference sequence and the subject sequence;

converting information of the extracted difference between the reference sequence and the subject sequence into a string of predetermined characters; and

encoding the individual characters that make the string of the characters using predetermined conversion codes that correspond to the individual characters.